

David C. Wheelock is a senior economist at the Federal Reserve Bank of St. Louis. Paul W. Wilson is an associate professor of economics at the University of Texas at Austin. Heidi L. Beyer provided research assistance.

Evaluating the Efficiency of Commercial Banks: Does Our View of What Banks Do Matter?

David C. Wheelock and Paul W. Wilson

In the past 15 years, the banking industry has faced growing competition from other financial service firms and financial markets and, at the same time, has undergone substantial deregulation and change. Proponents of further deregulation, such as the removal of barriers to the commingling of commercial and investment banking, argue that such changes would enhance the efficiency and viability of American banks.

The impact of competitive and regulatory changes on banks can be judged by gross measures of performance, such as profitability and failure rates. Economists are also interested in how such changes affect the efficiency with which banks transform resources into various financial services. Inefficiency implies that resources are wasted, that is, that firms are producing less than the feasible level of output from the resources employed, or are using relatively costly combinations of resources to produce a particular mix of products or services. Thus, a goal of policymakers, as well as stockholders and managers, is to devise policies that improve the efficiency of commercial banks.

Unfortunately, economists do not agree upon the appropriate methodology for measuring the efficiency of banks. Several

estimation techniques have been proposed, each with advantages and disadvantages. The problem is complicated by the myriad of different services that commercial banks perform. Researchers deal with complex issues in measuring bank production: Is a deposit an input to the production process, or an output? Should outputs be measured in terms of the number of a bank's accounts, the number of transactions it processes or the dollar amounts of its loans or deposits? Perhaps not surprisingly, estimates of commercial bank inefficiency vary considerably across studies that use different techniques, conceptions of bank production and data samples.

This article investigates the sensitivity of efficiency measures to broadly different conceptions of how banks operate. We use a single-estimation technique and a common pool of banks to compare efficiency measures based on alternative views of bank production. We find substantial differences in mean efficiency across models and low, though statistically significant, correspondence in the rankings of banks by efficiency scores across models.

First, we discuss why measuring commercial bank efficiency is useful, some alternative measures of efficiency and techniques for estimating efficiency. A description of the approach we take, our data and our results follow.

WHY DO WE CARE ABOUT THE EFFICIENCY OF COMMERCIAL BANKS?

The performance of firms is often described in terms of their efficiency. The measured efficiency of a production unit (a firm or plant) is generically interpreted as the difference between its observed input and output levels and the corresponding optimal values. An output-oriented measure of efficiency compares observed output with the maximum output possible for given input levels. Alternatively, an input-oriented

efficiency measure compares the observed level of inputs with the minimum input that could produce the observed level of output. These are measures of technical efficiency, and as such ignore the behavioral goals of the firm.

Measures of allocative efficiency compare the observed mix of inputs or outputs with the optimal mix that would minimize cost, maximize profit or obtain any other behavioral goal. Allocative efficiency can be combined with technical efficiency to measure overall efficiency. In addition, measures of technical efficiency can be used to construct measures of scale efficiency, which involve comparison of observed and optimal scale, or size, of the firm. One can also measure scope efficiency, which involves comparison of the cost of producing the observed mix of outputs in a single firm with the costs that would prevail if each output was produced in a separate firm. Researchers have found that banks suffer more from technical inefficiency than from scale or scope inefficiency (for example, Berger and Humphrey, 1991).

The efficiency of commercial banks is important for at least two reasons. First, efficiency measures are indicators of success, by which the performance of individual banks, and the industry as a whole, can be gauged. Banks face growing competition, both from other banks and from firms and markets outside the industry (see Wheelock, 1993), and presumably banks will be more successful in maintaining their business if they operate efficiently. Berger and Humphrey (1992) find that during the 1980s high-cost banks experienced higher rates of failure than more efficient banks. Similarly, in a study of bank failures during the 1920s, Wheelock and Wilson (1995) find that the less technically efficient a bank was, the greater its likelihood of failure.

A second reason to investigate the efficiency of commercial banks is the potential impact of government policies on efficiency. One might gauge the impact of a regulatory change by measuring its effect on commercial bank efficiency, or examine efficiency among banks in different states to measure the effect of differences in branching restric-

tions or other regulations. Recent proposals to end the Glass-Steagall separation of commercial and investment banking stem in part from a view that broader powers could enhance the efficiency of banks and other financial institutions. Obviously, this change could enhance the scope efficiency of banks if there are complementarities in the production of commercial and investment banking services. Conceivably, such change could also improve scale or overall efficiency. Improved efficiency is also one argument made in support of interstate branching and, indeed, Grabowski, Rangan and Rezvanian (1993) find that branch banking organizations are more efficient than multiple-office bank holding companies.

Other studies have considered whether bank mergers enhance efficiency. Using different approaches, Rhodes (1993) finds that mergers have not generally improved efficiency, though Fixler and Zieschang (1993) conclude the opposite. Shaffer (1993), on the other hand, evaluates potential mergers and concludes that they could significantly reduce inefficiency for many banks of less than \$10 billion of assets.

The impact of ownership or management structure on efficiency has also been studied. Pi and Timme (1993), for example, find that banks whose chief executive officer also serves as board chairman are less efficient than other banks, and Mester (1993) shows that mutual savings and loan associations are more efficient than stock S&Ls.

MEASURING COMMERCIAL BANK EFFICIENCY

The efficiency of commercial banks has been studied using a variety of techniques and samples, and, as noted above, has been used to address numerous policy issues. Recent studies typically use techniques that accommodate the multiple outputs of banks and measure the efficiency of individual banks relative to a standard set by peer institutions. Readers interested in a survey of this research can refer to Berger, Hunter and Timme (1993).

To date, no technique for measuring efficiency has been generally accepted and different methodologies appear to generate

considerable differences in measured efficiency, even when common bank samples are used. Variants of four techniques are common in the literature. The "stochastic cost frontier" approach is an econometric methodology in which deviations of a firm's actual cost from predicted cost are presumed to be due to random error and inefficiency, each of which is assumed to have a particular statistical distribution (usually the normal distribution for the random error and a half-normal for inefficiency). The "thick frontier" approach is a variant in which deviations from predicted cost within the lowest average cost quartile of banks are assumed due to random error, and the differences between the predicted costs of banks in the highest and lowest quartiles are assumed to be due to inefficiency. The "distribution-free" approach is applicable when data for more than one year are available. It assumes that inefficiency is stable over time, while random errors average out over time. That is, a bank's inefficiency for a span of years is taken to be the mean of its measured inefficiency across all years within the period. Finally, "Data Envelopment Analysis" (DEA) is a non-parametric methodology in which linear programming is used to measure the distance of individual banks from the efficient, or "best-practice," frontier. All deviations from the efficient frontier are assumed to be due to inefficiency.

Researchers have found that estimates of inefficiency are sensitive to the choice of technique. Ferrier and Lovell (1990), for example, apply the stochastic cost frontier and DEA techniques to a common sample of banks and arrive at different estimates of inefficiency. Berger (1993) finds substantial differences in measured efficiency from two variants of the distribution-free approach.

A second reason why different studies of commercial bank efficiency often reach seemingly contradictory findings might stem from differences in how a banking firm is modeled. Regardless of which of the four measurement techniques is used, the researcher must specify a list of inputs and outputs. The question, "What do banks produce?" is not simple to answer. Banks provide a variety of services, from loans

and deposit accounts to trust services, safe deposit box rentals, mutual fund sales and foreign exchange transactions. Moreover, changes in regulation, technology and customer demands have caused the types of services that banks perform to change over time. For example, banks now provide a variety of securities-related services, such as underwriting and mutual fund sales, which regulators forbid a few years ago. To tractably measure efficiency, researchers are forced to begin with simplified models of the banking firm. Unreliable estimates of efficiency can stem from the use of models that omit key features of bank production.

Some researchers view banks as producers of loans and deposit accounts, and measure output by either the number of transactions or accounts serviced. This view is referred to as the "production" approach. Others argue that a bank's output should be measured in terms of the dollar volume of loans or deposits it provides, a view known as the "intermediation" approach. Most studies of inefficiency use the intermediation approach, in part because the necessary data are more readily obtained. We are aware of only one recent study taking the production approach (Ferrier and Lovell, 1990), though in the 1970s and early 1980s such studies were more common (see Gilbert, 1984). The production approach focuses on operating costs and ignores interest expense. The intermediation approach, on the other hand, includes both operating and interest expenses, and hence may be of more interest for studying the viability of banks (see Berger, Hanweck and Humphrey, 1987; or Ferrier and Lovell, 1990). For analysis of the operating efficiency of banks, however, the production approach may be of interest.

Among those who use the intermediation approach are researchers who hold the view that banks produce various loans and other investments from deposits, other funding sources, labor and materials. This "asset" approach has been criticized because it ignores the fact that banks expend considerable resources supplying transactions and savings deposits (Berger and Humphrey, 1992).

Some researchers apply empirical criteria to determine what services to consider as bank outputs and what to consider as inputs.

Berger and Humphrey (1992), for example, classify activities for which banks create high added value, such as loans, demand deposits and time and savings deposits as important outputs, with labor, physical capital and purchased funds classified as inputs. Alternatively, Aly, Grabowski, Pasurka and Rangan (1990), Hancock (1991) and Fixler and Zieschang (1993) adopt a "user-cost" framework, whereby a bank asset is classified as an output if the financial return on the asset exceeds the opportunity cost of the investment, and a liability is classified as an output if the financial cost of the liability is less than its opportunity cost. Even though their details differ, the two approaches empirically tend to suggest similar classifications of inputs and outputs. The main exception is classification of demand deposits as an output in most user-cost studies, and as both an input and an output when the value-added approach is used (see Berger and Humphrey, 1992, for more detail).

Table 1 summarizes six recent studies of commercial bank production efficiency. Although representative, this list is far from exhaustive. These studies employ a variety of estimation techniques and include a variety of different inputs and outputs in modeling the banking firm. The studies typically report inefficiency measures by bank-size grouping and for more than one type of inefficiency, though for brevity we report just the mean overall inefficiency. The reported percentages indicate the extent to which the average bank overused inputs to produce a given level of output. Thus, the 35 percent inefficiency found by Aly and others (1990) indicates that the average bank could have produced the same level of output with just 65 percent of the input levels actually used. Measured inefficiency clearly varies with estimation technique, model specification and, probably, the sample of banks used by the researcher.

In the remainder of this article, we investigate the extent to which measures of efficiency and the rankings of individual banks depend on whether the intermediation approach or production approach is employed. Because we are interested in the impact of the approach taken on measured efficiency, we use a single technique—DEA—applied

to a common pool of banks. Our findings might, of course, be different if we used another technique or sample, but the purpose of this article is to investigate how sensitive efficiency measures are to the model of bank production employed.

METHODOLOGY

We trace our measures of efficiency to the work of Debreu (1951) and Farrell (1957). Boles (1966) was one of the first to use linear programming methods to measure efficiency in production using their ideas. Other extensions have collectively come to be named Data Envelopment Analysis (DEA), a term coined by Charnes, Cooper and Rhodes (1978). Lovell (1993) summarizes this literature.

Details about the efficiency measures used in this article are contained in the shaded insert on page 6. The essential ideas, however, are illustrated in Figure 1, which considers the case of a sample of firms producing a single output from two inputs, x_1 and x_2 . Suppose firms A, B and C each produce a given level of output; A and B lie on the production frontier XX' , while C lies in the interior of the production set. The frontier XX' is the set of all combinations of inputs which can produce the same level of output, and where the reduction of at least one input necessarily causes output to fall. Hence, firms A and B are regarded as efficient, whereas firm C is regarded as inefficient. Inefficient firms such as C may lie in the interior of the production set due to imperfect information, managerial incompetence or perhaps other reasons. For firm C, input weak technical efficiency (IWE) is defined as the ratio of distances OC'/OC in Figure 1. By reducing the input quantities used by firm C by this amount, the firm could move to point C' and would be considered efficient in the IWE sense.

Next, we define input overall efficiency (IOE). In terms of Figure 1, the isocost line is given by PP' . For firm C, the IOE score is given by the ratio of distances OC''/OC . Although the point C'' lies outside the production set boundary, and hence is not feasible, input costs at C'' are the same as at B, which is a feasible point. Hence, if firm C

Table 1

Selected Studies of Commercial Bank Production Inefficiency

Study, Technique, Approach	Inputs	Outputs	Sample	Results
Aly and others (1990); DEA; intermediation	labor, physical capital, loanable funds ¹	real estate loans, com- mercial loans, consumer loans, all other loans, demand deposits	random, 322 banks, 1986 data	overall inefficiency: 35%
Berger and Humphrey (1991); thick frontier; intermediation	labor, physical capital, purchased funds ²	demand deposits, retail time and savings deposits, real estate loans, commercial loans, installment loans	all banks, 1984 data	total inefficiency: ³ 24% (branching states), 19% (unit banking states)
Elyasiani and Mehdi- an (1990); DEA intermediation	labor, physical capital, demand deposits, time and savings deposits	securities held, real estate loans, commercial loans, all other loans	191 banks with assets over \$300 million, 1980 data	technical inefficiency: 10%
Ferrier and Lovell (1990) DEA and stochastic cost frontier; production	labor, occupancy costs, expenditure on material	number of: demand deposit accounts, time and savings deposit accounts, real estate loans, installment loans, commercial loans	575 banks, 1984 data	overall inefficiency: 21% (DEA), 26% (stochastic cost frontier)
Hunter and Timme (1995) distribution free; intermediation	labor, physical capital, ⁴ purchased funds, transactions accounts, ⁴ non-trans- actions accounts under \$100,000 ⁴	commercial and security loans, consumer loans, all other loans, non- interest income	317 banks with assets over \$1 bil- lion, 1985-90 data	overall inefficiency: 30-54% (depend- ing on model); 23- 36% (omitting 1% extreme values)
Kaparakis and others (1994); stochastic cost frontier; intermediation	labor, physical capital, interest bearing deposits under \$100,000, non-inter- est bearing deposits, ⁴ purchased funds	loans to individuals, real estate loans, commercial loans, other ⁵	5,548 banks with assets over \$50 million, 1986 data	overall inefficiency: 10%

¹ time and savings deposits, notes and debentures and other borrowed funds.

² federal funds purchased, time deposits over \$100,000, foreign deposits and other borrowed funds.

³ includes inefficiencies due to excessive deposit interest paid and purchased fund interest paid.

⁴ input treated as "quasi-fixed," that is, not variable in the short run.

⁵ fed funds sold, securities held, securities and other assets in trading accounts.

were to become efficient in the IOE sense, its input mix would have to be altered; the IOE score, however, can be obtained by considering the hypothetical proportionate reductions of inputs represented by point C".

In terms of Figure 1, allocative efficiency for firm C is given by the ratio of distances OC"/OC'. Allocative inefficiency arises from using a combination of inputs that does not

minimize total cost, as opposed to technical inefficiency, which is a proportionate overuse of all inputs.

Finally, we can determine scale efficiency by comparing IWE computed under the assumption that the firm is operating at constant returns-to-scale with IWE obtained previously. A score of unity implies that the firm is operating under constant returns. While a score other than 1 does not translate

A MATHEMATICAL DESCRIPTION OF EFFICIENCY MEASUREMENT

We use measures of efficiency discussed by Fare, Grosskopf and Lovell (1985). First, we compute the input weak technical efficiency (IWE) score for the i th firm in a sample by solving the linear programming problem:

$$(1) \quad \begin{aligned} &\min W_i \\ &\text{subject to} \quad Xq_i \leq W_i x_i \\ &\quad Yq_i \geq y_i \\ &\quad \bar{1}q_i = 1 \\ &\quad q_i \in \mathcal{R}_+^N, \end{aligned}$$

where n firms produce s outputs using m inputs, q_i is a $(N \times 1)$ vector of weights to be computed for the i th firm, $0 < W_i \leq 1$ is a scalar, x_i is a $(m \times 1)$ vector of inputs for the k th firm, y_i is a $(s \times 1)$ vector of outputs for the k th firm, $X = [x_1, \dots, x_N]$ is a $(m \times N)$ matrix of observed inputs, $Y = [y_1, \dots, y_N]$ is a $(s \times N)$ matrix of observed outputs and $\bar{1}$ is a $(1 \times N)$ vector of ones.

The minimand W_i in equation 1 measures the input weak efficiency of the i th firm. The inequality constraints in equation 1 define a reference technology with strong disposability of inputs; constraining the weights in q to sum to unity allows the reference technology to exhibit variable returns to scale. For the i th firm, W_i gives the proportion by which inputs can be reduced to move the firm from the interior of the production set onto the piecewise-linear boundary of the production set corresponding to the reference technology in 1.

Next, we compute input overall efficiency (IOE) score O_i for the i th firm by first solving the linear program:

$$(2) \quad \begin{aligned} &\min_{x_i^*} p_i x_i^* \\ &\text{subject to} \quad Xq_i \leq x_i^* \\ &\quad Yq_i \geq y_i \\ &\quad \bar{1}q_i = 1 \\ &\quad q_i \in \mathcal{R}_+^N \\ &\quad x_i^* \in \mathcal{R}_+^m \end{aligned}$$

where X , Y , x_i and y_i are defined as in equation 1, p_i is a $(1 \times m)$ vector of input prices, and x_i^* is an $(m \times 1)$ vector of efficient inputs to be computed. The IOE score may be defined as

$$(3) \quad O_i \equiv p_i x_i^* / p_i x_i.$$

The constraints in equation 2 are similar to those in 1. The same reference technology is defined by the constraints in 2, but instead of proportionately reducing inputs until the i th firm lies on the reference technology, inputs are further reduced proportionately until the firm lies on the isocost plane tangent to the production set boundary.

An allocative efficiency score, A_i , may be defined by dividing the IOE score by the IWE score:

$$(4) \quad A_i \equiv O_i / W_i.$$

The efficiency scores obtained from 1 measure technical efficiency as the distance to the relevant isoquant, but do not consider where the firm is situated along the variable-returns production frontier. To measure scale efficiency, equation 1 must be recomputed for each firm, first assuming constant returns to scale by removing the restriction $Cq_i = 1$, and then assuming non-increasing returns-to-scale by imposing the restriction $Cq_i \leq 1$. In the case of IWE, this produces efficiency scores W_i^{CRS} and W_i^{NIRS} , respectively, for the i th firm. The scale efficiency score corresponding to 1 is then defined as

$$(5) \quad S_i \equiv W_i^{CRS} / W_i.$$

Clearly, $0 < S_i \leq 1$ since $W_i^{CRS} \leq W_i^{NIRS} \leq W_i$. If $S_i = 1$, then the i th firm is scale-efficient, that is, the firm is operating at the point of constant returns on the production frontier. If $S_i < 1$, then the firm is scale-inefficient due to either decreasing returns if $W_i^{NIRS} = W_i$, or increasing returns if $W_i^{NIRS} < W_i$.

easily into a specific percentage deviation from constant returns, the scores are useful for ranking firms by the extent of their inefficiency.

Each of the efficiency scores described above measures efficiency in an input orientation; efficiency is measured by holding output fixed and determining the maximum feasible reduction in inputs. Efficiency can also be measured by holding inputs fixed and determining the maximum feasible expansion of outputs. Since the efficiency measures we use do not imply underlying assumptions regarding the behavior of firms, the choice between input and output orientations is somewhat arbitrary; one might compute both types of efficiency measures to get more information than can be obtained from either the input or output orientations alone. Note that both IWE and IOE are *radial* measures of efficiency, that is, in each case efficiency is measured along a ray emanating from the origin and passing through the firm in input-output space. Consequently, the efficiency scores are independent of the units of measurement used for both inputs and outputs, which is advantageous since units of measurement may always be defined arbitrarily. Fare and others (1985) observe that some DEA formulations do not share this property.

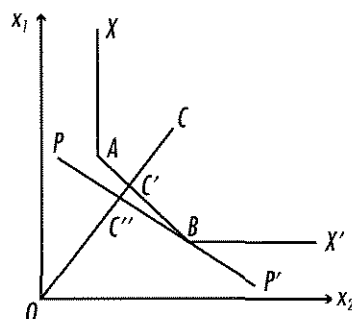
EMPIRICAL IMPLEMENTATION

For our empirical analysis of commercial bank efficiency, we use a sample of banks participating in the Federal Reserve System's Functional Cost Analysis (FCA) program for 1993. Participants in this program supply information about their operations and costs which are not generally available for banks, and which are necessary to measure efficiency using the production approach. After eliminating observations with missing values and observations for depository institutions other than commercial banks, data for 269 banks remain.

Because participation in the FCA program is voluntary, the banks in our sample may not be representative of the industry as a whole. For example, whereas the aver-

Figure 1

Measuring Technical, Allocative and Overall Efficiency



age total assets at the end of 1993 for FCA program banks was \$163.6 million, with a range from \$8.0 million to \$2,602.8 million, average total assets were \$300.7 million, with a range from \$1.0 million to \$108,223.0 million, for all U.S. commercial banks (as reported in the Federal Deposit Insurance Corporation Reports of Condition, that is, the "Call Reports"). The average return on assets of 1.15 percent for the banks in our sample, however, was approximately the same as the average for all banks (1.12 percent).

Nevertheless, because our sample of banks is not random, the efficiency measures calculated here should not be interpreted as reflecting the efficiency of commercial banks in general.

For the production approach to modeling bank activities, we construct variables using definitions from Ferrier and Lovell (1990):

Outputs:

- y_1 = number of demand deposit accounts
- y_2 = number of time deposit accounts
- y_3 = number of real estate loans
- y_4 = number of installment loans
- y_5 = number of commercial loans

Inputs:

- x_1 = number of employees
- x_2 = occupancy costs and expenditure on furniture and equipment
- x_3 = expenditure on materials

Table 2

Descriptive Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
y_1	7902.91	13091.76	469.00	173362.00
y_2	7618.43	10680.64	413.00	106821.00
y_3	998.23	1503.99	0.00	13456.00
y_4	3134.01	6918.63	44.00	87794.00
y_5	899.63	1744.37	0.00	23998.00
y_1	40764316.60	60020700.09	1387963.00	653519000.00
y_2	99272775.89	144141890.79	5103000.00	1616691000.00
y_3	42038331.61	54512376.38	0.00	377449000.00
y_4	16254832.77	43838042.24	127890.00	617136000.00
y_5	28553775.52	64197194.17	0.00	895471000.00
x_1	88.02	144.09	3.89	1730.07
x_2	776304.50	1260920.37	790.00	13090834.00
x_3	330106.61	576912.53	11859.00	6948552.00
w_1	30827.13	5905.65	19222.38	67832.63
w_2	0.0054	0.0025	0.0001	0.0270
w_3	0.0023	0.0007	0.0010	0.0057
w_1	30827.13	5905.65	19222.38	67832.63
w_2	52.06	48.43	0.75	715.17
w_3	21.88	11.69	2.36	131.86
v_1	17045.20	37140.71	166.00	431227.00
v_2	54417.53	96144.16	213.00	1279962.00
v_3	16755.40	29282.60	93.00	270480.00
v_4	5980.03	12048.51	100.00	137300.00
u_1	110490.98	158765.47	5138.00	1894477.00
u_2	18323.86	43312.12	200.00	410249.00
u_3	89.32	150.08	4.00	1856.00
u_4	3008.39	5252.85	13.00	47511.00
p_1	0.0347	0.0048	0.0143	0.0489
p_2	0.0392	0.0110	0.0066	0.1000
p_3	31.11	6.18	20.38	77.91
p_4	0.34	0.31	0.08	2.83

Input prices:

- w_1 = total expenditure on salaries and fringe benefits/ x_1
- w_2 = x_2 /level of deposits
- w_3 = x_3 /level of deposits

For the intermediation approach to bank production, we construct variables using definitions from Kaparakis, Miller and Noulas (1994):

Outputs:

- v_1 = loans to individuals for household, family, and other personal expenses
- v_2 = real estate loans
- v_3 = commercial and industrial loans
- v_4 = federal funds sold, securities purchased under agreements to resell, plus total securities held in trading accounts

Inputs:

- u_1 = interest-bearing deposits except certificates of deposit greater than \$100,000
- u_2 = purchased funds (certificates of deposit greater than \$100,000, federal funds purchased, and securities sold plus demand notes) and other borrowed money
- u_3 = number of employees
- u_4 = premises and fixed assets

Input prices:

- p_1 = average interest cost per dollar of u_1
- p_2 = average interest cost per dollar of u_2
- p_3 = average annual wage per employee
- p_4 = average cost of premises and fixed assets.

In addition, Kaparakis and others also define quasi-fixed input, non-interest bearing deposits, for which there is no corresponding price. Other studies adopting the intermediation approach have ignored this item, as we do in the results reported below. Including non-interest bearing deposits as a fifth input when measuring technical or scale efficiency seems to have little effect on the results.

To form a specification midway between the production approach represented by the Ferrier and Lovell (1990) specification and the intermediation approach represented by the Kaparakis and others (1994) specification, we define y'_1, \dots, y'_5 as the dollar amount of each account or loan corresponding to y_1, \dots, y_5 , respectively. Because outputs are now measured in dollar amounts, this model is best classified as representing the intermediation approach, even though the choice of variables is based on Ferrier and Lovell (1990). In addition, we define an alternative price system, w'_1, w'_2, w'_3 , for the Ferrier and Lovell specification, where $w'_1 = w_1$, and w'_2 and w'_3 are computed similarly to w_2 and w_3 except that level of deposits is replaced by the *number* of time and demand deposits. This seems to us to make the mapping of inputs and outputs under the production approach more consistent. We report summary statistics for each of the variables in Table 2.

RESULTS OF EFFICIENCY MEASUREMENT

We compute the various efficiency measures for the five models summarized below:

Model	Inputs	Outputs	Input Prices
1	$x_1 - x_3$	$y_1 - y_5$	$w_1 - w_2$
2	$x_1 - x_3$	$y_1 - y_5$	$w'_1 - w'_2$
3	$x_1 - x_3$	$y'_1 - y'_5$	$w_1 - w_2$
4	$x_1 - x_3$	$y'_1 - y'_5$	$w'_1 - w'_2$
5	$u_1 - u_4$	$v_1 - v_4$	$p_1 - p_4$

Models 1 and 2 correspond to the Ferrier and Lovell (1990) specification, with alternative price definitions. Models 3 and 4 provide a bridge to the intermediation approach by replacing the number of accounts and loans in the output variables with dollar amounts. Model 5 is the Kaparakis and others (1994) specification.

Table 3 presents the mean scores for each type of efficiency described in the preceding section. Note that since the same inputs and outputs are used in models 1 and 2, and models 3 and 4, the technical and scale efficiency scores are the same for these models. For each efficiency measure, Table 3 also shows the standard deviation of the scores across the 269 banks in the sample, the number of banks having an efficiency score of unity (labeled "Number Efficient"), that is, the number of banks operating on the efficient frontier, as well as 90 and 95 percent confidence intervals for the mean. Given the large number of banks with efficiency scores of unity, particularly in the case of technical efficiency, and since all of the efficiency scores are defined to lie between zero and 1, the underlying distributions of the individual efficiency scores are clearly non-normal. Results from Atkinson and Wilson (1995), however, suggest that our sample size of 269 is easily large enough for us to rely on the asymptotic normality of the sample means implied by the central limit theorem, and thus to compute confidence intervals based on a normal distribution.

In several cases, the confidence intervals for the means reported in Table 3 overlap.

Table 3

Efficiency Scores (269 observations)

Model	Mean	Standard Error	Number Efficient	90% Confidence Interval		95% Confidence Interval	
Technical efficiency (W_k):							
1,2	0.6348	0.0141	42	0.6115	0.6581	0.6071	0.6626
3,4	0.7675	0.0103	52	0.7505	0.7846	0.7472	0.7879
5	0.8088	0.0107	75	0.7911	0.8265	0.7877	0.8299
Scale efficiency (S_k):							
1,2	0.8833	0.0067	32	0.8723	0.8943	0.8701	0.8964
3,4	0.9452	0.0047	49	0.9374	0.9530	0.9359	0.9545
5	0.9414	0.0057	7	0.9319	0.9509	0.9301	0.9527
Allocative efficiency (A_k):							
1	0.7698	0.0082	7	0.7562	0.7833	0.7536	0.7860
2	0.4992	0.0108	4	0.4814	0.5170	0.4780	0.5204
3	0.7924	0.0068	7	0.7812	0.8036	0.7790	0.8057
4	0.6340	0.0108	7	0.6162	0.6518	0.6128	0.6552
5	0.7838	0.0080	13	0.7706	0.7970	0.7680	0.7996
Overall efficiency (O_k):							
1	0.4835	0.0116	7	0.4644	0.5026	0.4607	0.5063
2	0.3356	0.0128	4	0.3146	0.3567	0.3105	0.3607
3	0.6053	0.0095	7	0.5897	0.6209	0.5867	0.6239
4	0.4928	0.0121	7	0.4727	0.5128	0.4689	0.5166
5	0.6320	0.0107	13	0.6144	0.6496	0.6110	0.6529

We test for significant differences among the means of each efficiency measure across different models. At the 0.05 significance level, we are unable to reject the null hypothesis of equivalent means in the following cases: (1) scale efficiency for models 3, 4 and 5; (2) allocative efficiency for models 1 and 5, and models 3 and 5 (we do reject the null hypothesis when comparing allocative efficiency among models 1 and 3); and (3) overall efficiency for models 1 and 4. In all other instances, we reject the null hypothesis of no difference. Even the seemingly innocuous modification of redefining the input prices between models 1 and 2, and between 3 and 4 has a large effect on mean allocative and overall efficiency. Note also that for the most extreme comparison, models 1 (the production view) and 5

(the intermediation view), we reject the null hypothesis of equal levels of technical and overall efficiency. This suggests that, at least for this sample of banks, average efficiency does depend on the view of bank production assumed by the researcher. We find that average technical and overall efficiency is higher under the intermediation approach (model 5) than under the production approach (model 1). Our finding for overall inefficiency of 37 percent using model 5 is similar to what Aly and others (1990) found for their sample, though substantially greater than what Kaparakis and others (1994) found for theirs (see Table 1).

It is possible to determine whether a particular bank lies on the increasing (IRS), constant (CRS) or decreasing (DRS) returns portion of the technology. Table 4 shows

the results of this analysis, considering only banks that were found to be technically efficient.¹ Thus, for example, 16 banks, or 38.1 percent of all technically efficient banks, operated on the constant-returns portion of the technology under models 1 and 2.

We test the null hypothesis of no association among the rows and columns of the matrix represented by Table 4 using Pearson's chi-square test, the likelihood ratio chi-square statistic, and Fisher's exact test.² For the entire matrix, all three tests reject the null hypothesis of no differences in the proportions in each row and column. However, when we perform pairwise tests by deleting individual rows from Table 4, we fail to reject the null hypothesis of no difference for models 3, 4 and 5, and for models 1, 2 and 3, 4. Each of the three tests fail to reject at the 90 percent level.

In the case of models 1, 2 and 5, we reject the null hypothesis of no difference in the proportions at greater than 99 percent. Thus, while we find evidence of similarity in terms of returns-to-scale when comparing models 3, 4 with either models 1, 2 or 5, models 1, 2 and 5 appear different in terms of returns-to-scale. More banks appear to be operating under constant returns-to-scale when the intermediation approach is taken (model 5) than when the production approach is used (model 1). Since returns-to-scale at a given location on the production frontier depend upon the shape of the variable-returns technology, these results indicate that the technology implicitly estimated by models 3, 4 is similar to the technologies implied by models 1, 2 and 5, which in turn are significantly different. This is consistent with our view of models 3, 4 as a bridge between the production approach represented by models 1, 2 and the intermediation approach represented by model 5. The result also suggests that differences between the two approaches might be due not only to use of number of accounts and loans versus dollar amounts, but also to the treatment of time deposits as an output or an input.

In addition to comparing mean efficiency scores, we use the Wilcoxon matched-pairs signed-ranks test, a sign test for equality of medians and Kendall's τ -statistic to further

Table 4

Returns to Scale

Model	CRS	DRS	IRS
1,2	16 (38.1%)	24 (57.1%)	2 (4.8%)
3,4	31 (59.6%)	20 (38.5%)	1 (1.9%)
5	48 (64.0%)	21 (28.0%)	6 (8.0%)

examine the similarity of efficiency scores across the five models. We report the results of these tests in Table 5.³

The Wilcoxon matched-pairs signed-ranks test analyzes the equality of distributions without making assumptions regarding the form the distributions might take. The values shown in parentheses in the second column of Table 5 give the two-tailed normal probabilities associated with the test statistic. Hence, we fail to reject the null hypothesis of identical distributions when comparing scale efficiency scores from models 3, 4 and 5, when comparing allocative efficiency scores from models 1 and 5, and from models 3 and 5, and when comparing overall efficiency scores from models 1 and 4. In all other cases, we reject the null hypothesis. It appears that, for the most part, the distributions of the various efficiency scores do vary across models.

The sign test for equivalence of medians yields a two-tailed binomial probability, which we also report in Table 5. In only two instances do we not reject the null hypothesis of equal medians: when comparing scale efficiency scores from models 3, 4 and 5, and when comparing allocative efficiency scores from models 3 and 5. These results are consistent with our finding that, in most cases, average efficiency varies across models.

Finally, rather than comparing the distributions of efficiency scores from different models, we use the scores to *rank* banks in terms of their estimated efficiency. Kendall's τ -statistic measures the correlation among the ranks of banks from two models and provides a statistical test of the null hypothesis of no association between two sets of

¹ Since we are using an input orientation, we could also examine whether inefficient banks would lie on the increasing-, constant- or decreasing-returns portion of the technology if inputs were proportionately contracted to move the bank to the frontier. However, since the path a bank might actually take to reach the frontier if the sources of inefficiency were removed depends upon behavioral goals, we ignore technically inefficient banks here.

² Details on these computations may be found in the *Stata Reference Manual: Release 3.1*, Stata Corporation (1993).

³ See Snedecor and Cochran (1989) and Kendall and Gibbons (1990) for a discussion of these tests. Computational details are given in the *Stata Reference Manual: Release 3.1*, Stata Corporation (1993).

Table 5

Statistical Comparison of Efficiency Scores
(probability values in parentheses)

Model	Wilcoxon Signed-Ranks	Sign Test	Kendall's τ
Technical efficiency (W_x):			
(2,2) / (3,4)	-8.24 (0.0001)	0.0000	0.2513 (0.0000)
(1,2) / 5	-8.90 (0.0001)	0.0000	0.0909 (0.0262)
(3,4) / 5	-3.25 (0.0012)	0.0034	0.1081 (0.0082)
Scale efficiency (S_k):			
(1,2) / (3,4)	-9.46 (0.0001)	0.0000	0.2333 (0.0000)
(1,2) / 5	-7.58 (0.0001)	0.0000	0.1114 (0.0065)
(3,4) / 5	-0.32 (0.7482)	0.4644	0.1805 (0.0000)
Allocative efficiency (A_k):			
(1,2)	12.88 (0.0001)	0.0000	0.0068 (0.8680)
(1,3)	-3.58 (0.0003)	0.0003	0.4179 (0.0000)
(1,4)	8.40 (0.0001)	0.0000	-0.0964 (0.0185)
(1,5)	-1.45 (0.1481)	0.0327	0.0934 (0.0225)
(2,3)	-13.39 (0.0001)	0.0000	0.0267 (0.5135)
(2,4)	-12.28 (0.0001)	0.0000	0.5661 (0.0000)
(2,5)	-13.03 (0.0001)	0.0000	0.0342 (0.4036)
(3,4)	10.23 (0.0001)	0.0000	-0.0047 (0.9092)
(3,5)	0.32 (0.7515)	0.6258	0.0627 (0.1255)
(4,5)	-10.49 (0.0001)	0.0000	0.1648 (0.0001)

Table 5 (cont.)

Statistical Comparison of Efficiency Scores
(probability values in parentheses)

Model	Wilcoxon Signed-Ranks	Sign Test	Kendall's τ
Overall efficiency (O_k):			
(1,2)	12.31 (0.0001)	0.0000	0.5874 (0.0000)
(1,3)	-9.08 (0.0001)	0.0000	0.2518 (0.0000)
(1,4)	-1.06 (0.2903)	0.0327	0.3072 (0.0000)
(1,5)	-9.17 (0.0001)	0.0000	0.1329 (0.0012)
(2,3)	-12.74 (0.0001)	0.0000	0.1704 (0.0000)
(2,4)	-11.62 (0.0001)	0.0000	0.5000 (0.0000)
(2,5)	-12.36 (0.0001)	0.0000	0.0405 (0.3222)
(3,4)	9.76 (0.0001)	0.0000	0.4128 (0.0000)
(3,5)	-2.67 (0.0076)	0.0015	0.1034 (0.0115)
(4,5)	-9.01 (0.0001)	0.0000	0.1831 (0.0000)

rankings. The statistic is approximately normally distributed, with zero expected value and with variance

$$\text{VAR}(\tau) = \frac{4N + 10}{9N(N - 1)},$$

where N gives the number of observations. By definition, the statistic lies between -1 and +1, taking a value of +1 if rankings are in complete agreement, or -1 if the ranks are completely reversed.

Kendall and Gibbons (1990) suggest that the τ -statistic may also be viewed as a measure of concordance. Any two pairs of ranks (u_i, v_i) and (u_j, v_j) , $i, j = 1, \dots, N$, $i \neq j$, are defined as concordant if $v_i < v_j$ when $u_i < u_j$ or $v_i > v_j$ when $u_i > u_j$. Similarly, they are defined as discordant if $v_i < v_j$ when $u_i > u_j$ or $v_i > v_j$ when $u_i < u_j$. The total number of pairs is

$N(N-1)/2$, and τ can be shown to be equivalent to the proportion of concordant pairs minus the proportion of discordant pairs.

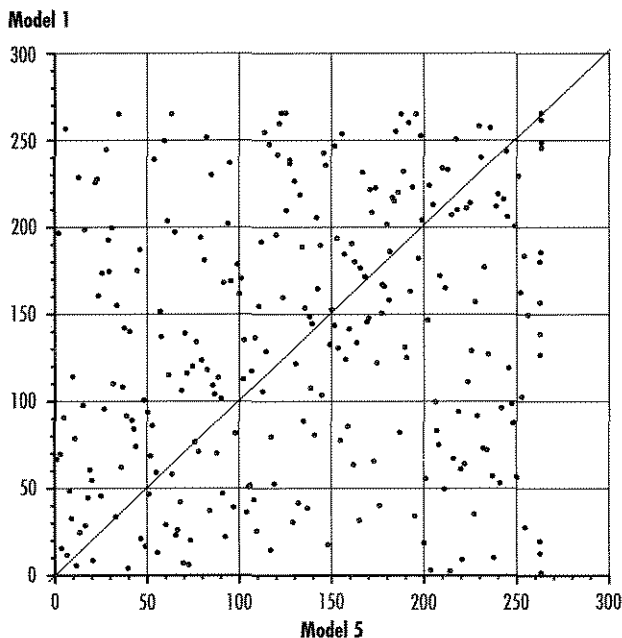
The last column of Table 5 gives the τ -statistic for the various pairs of models for each measure of efficiency, along with significance levels as shown in parentheses. We fail to reject the null hypothesis of no association among ranks in only five instances when comparing allocative efficiency scores. (In particular, we fail to reject the null hypothesis for the following pairs of models: 1, 2; 2, 3; 2, 5; 3, 4; and 3, 5, and in only one case when comparing overall efficiency scores—for models 2 and 5). In all other cases, we reject the null hypothesis of no correlation.

Note, however, that when we reject the null hypothesis of no association, the τ -statistic is usually rather small in absolute terms; the largest value of the statistic shown in Table 5 is 0.5874 (in the case of overall efficiency for models 1 and 2). As is typical in classical hypothesis testing, rejection of the null hypothesis does not necessarily imply acceptance of an alternative hypothesis. That is, our statistical test may reject the hypothesis that the rankings are not associated, but that does not necessarily imply that the rankings are associated—the test is simply not that powerful. Figure 2 plots the rankings of overall efficiency scores for model 1 against those for model 5. Note that the value of Kendall's τ -statistic for this comparison is significantly different from zero at the 0.0012 level, indicating that the two sets of rankings are not discordant. The low value of the test statistic (0.1329), however, suggests that neither are they concordant.

Our results based on the Wilcoxon matched-pairs signed-ranks test and the sign test for equivalence of medians are consistent with our observations on the differences of mean efficiency scores across the models discussed earlier. Taken together, our results indicate that different model specifications are likely to produce different measures of the level of inefficiency among a sample of banks, but not necessarily dissimilar rankings of individual banks in terms of measured efficiency. For our data, the rankings are similar enough to reject the

Figure 2

Rankings of Overall Efficiency Scores



null hypothesis of no association, but in many cases are far from being in complete concordance. Concordance is relatively high for the technical and scale efficiency measures, which do not rely on price data. The introduction of price data to measure allocative and overall efficiency might also introduce more sources of noise or error.

CONCLUSION

Like other studies of commercial bank efficiency, we find considerable inefficiency among banks in our sample. Other studies have found substantial variation in efficiency measures in applying different estimation techniques to a common pool of banks. We find that measured efficiency also depends on the researcher's conception of what banks do. In this article, we measure various types of production efficiency under two very different views of banking. We find that, on average, technical and overall efficiency is higher under the intermediation view of the banking firm than under the production view. Mean allocative efficiency is, however, similar

under the extreme versions of each approach. Under the intermediation view, we also find somewhat less scale inefficiency and more banks operating on the constant-returns portion of the efficient frontier. Despite the differences in mean measured efficiency across the different conceptions of how banks operate, however, we find some similarity in the rankings of efficiency scores of individual banks. Further research will, of course, be necessary to determine how sensitive these findings are to the particular dataset and estimation techniques employed in this article.

REFERENCES

- Aly, Hassan Y., Richard Grabowski, Carl Pasurka and Nanda Rangan. "Technical, Scale, and Allocative Efficiencies in U.S. Banking: An Empirical Investigation," *Review of Economics and Statistics* (May 1990), pp. 211-18.
- Atkinson, Scott, and Paul W. Wilson. "Comparing Mean Efficiency and Productivity Scores Across Small Samples: A Bootstrap Methodology," *Journal of Productivity Analysis* (forthcoming).
- Berger, Allen N. "Distribution-Free Estimates of Efficiency in the U.S. Banking Industry and Tests of the Standard Distributional Assumptions," *Journal of Productivity Analysis* (1993), pp. 261-92.
- _____, G. A. Honweck and D. B. Humphrey. "Competitive Viability in Banking: Scale, Scope, and Product Mix Economies," *Journal of Monetary Economics* (1987), pp. 501-20.
- _____, and David B. Humphrey. "Measurement and Efficiency Issues in Commercial Banking," in Zvi Griliches, ed., *Output Measurement in the Services Sector*. University of Chicago Press, 1992, pp. 245-79.
- _____, and _____. "The Dominance of Inefficiencies Over Scale and Product Mix Economies in Banking," *Journal of Monetary Economics* (August 1991), pp. 117-48.
- _____, William C. Hunter and Stephen G. Timme. "The Efficiency of Financial Institutions: A Review and Preview of Research, Past, Present and Future," *Journal of Banking and Finance* (April 1993), pp. 221-49.
- Boles, J. "Efficiency Squared: Efficient Computations of Efficiency Indexes," *Western Farm Economics Association Proceedings* (1966), pp. 137-42.
- Charnes, A., W. W. Cooper and E. Rhodes. "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research* (1978), pp. 429-44.
- Debreu, Gerard. "The Coefficient of Resource Utilization," *Econometrica* (1951), pp. 273-92.
- Elyasiani, Elyas, and Seyed M. Mehdi. "A Nonparametric Approach to Measurement of Efficiency and Technological Change: The Case of Large U.S. Commercial Banks," *Journal of Financial Services Research* (July 1990), pp. 157-68.
- Färe, Rolf, Shawna Grosskopf and C. A. Knox Lovell. *The Measurement of Efficiency of Production*. Kluwer-Nijhoff Publishing, 1985.
- Farrell, M. J. "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society, Series A* (1957), pp. 253-81.
- Ferrier, Gary D., and C. A. Knox Lovell. "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence," *Journal of Econometrics* (October/November 1990), pp. 229-45.
- Fixler, Dennis J., and Kimberly D. Zieschang. "An Index Number Approach to Measuring Bank Efficiency: An Application to Mergers," *Journal of Banking and Finance* (April 1993), pp. 437-50.
- Gilbert, R. Alton. "Bank Market Structure and Competition: A Survey," *Journal of Money, Credit and Banking* (November 1984), pp. 617-45.
- Grabowski, Richard, Nanda Rangan and Rasoul Rezvanian. "Organizational Forms in Banking: An Empirical Investigation of Cost Efficiency," *Journal of Banking and Finance* (April 1993), pp. 531-38.
- Hancock, Diana. *A Theory of Production for the Financial Firm*. Kluwer Academic Publishers, 1991.
- Hunter, William C., and Stephen G. Timme. "Core Deposits and Physical Capital: A Reexamination of Bank Scale Economies and Efficiency with Quasi-Fixed Inputs," *Journal of Money, Credit and Banking* (February 1995), pp. 165-85.
- Karakakis, Emmanuel I., Stephen M. Miller and Athanasios G. Noulas. "Short-run Cost Inefficiency of Commercial Banks: A Flexible Stochastic Frontier Approach," *Journal of Money, Credit and Banking* (November 1994), pp. 875-93.
- Kendall, M., and J. D. Gibbons. *Rank Correlation Methods*. Oxford University Press, 1990.
- Lovell, C. A. Knox. "Production Frontiers and Productive Efficiency," in Hal Fried, C. A. Knox Lovell and Shelton S. Schmidt, eds., *The Measurement of Productive Efficiency: Techniques and Applications*. Oxford University Press, 1993, pp. 3-67.
- Mester, Loretta J. "Efficiency in the Savings and Loan Industry," *Journal of Banking and Finance* (April 1993), pp. 267-86.
- Pi, Lynn, and Stephen G. Timme. "Corporate Control and Bank Efficiency," *Journal of Banking and Finance* (April 1993), pp. 515-30.
- Rhoades, Stephen A. "Efficiency Effects of Horizontal (In-Market) Bank Mergers," *Journal of Banking and Finance* (April 1993), pp. 411-22.
- Shaffer, Sherrill. "Can Megamergers Improve Bank Efficiency?" *Journal of Banking and Finance* (April 1993), pp. 423-36.
- Snedecor, G. W., and W. G. Cochran. *Statistical Methods*. Iowa State University Press, 1989.
- Stata Reference Manual: Release 3.1*. Stata Corporation, 1993.
- Wheelock, David C. "Is the Banking Industry in Decline? Recent Trends and Future Prospects from a Historical Perspective," this *Review* (September/October 1993), pp. 3-22.
- _____, and Paul W. Wilson. "Explaining Bank Failures: Deposit Insurance, Regulation and Efficiency," *Review of Economics and Statistics* (forthcoming).